Semantic Space Segmentation for Content-Based Image Retrieval using SVM Decision Boundary and Principal Axis Analysis

Tian-Luu Wu and Ji-Hwei Horng

Abstract—This paper presents a semantic space segmentation method for content-based image retrieval using the SVM decision boundary and principal axis analysis. Each object in an image has its activity scope, which conforms to semantic concept invariance. To find the maximum and fair activity scope for every object in an image in the best way is to discriminate the hidden semantic differences. A special property of SVM is known as maximum margin classifier, which simultaneously minimizes the empirical classification error and maximize the geometric margin. In this paper, the hidden semantic concepts based on the low-level features of region (object) will be found by using the Expectation-Maximization (EM) technique. The activity scope of each object and support vector will be decided by the SVM decision boundary and principal axis analysis which is used to construct a hyper-plane projection line with minimum inertia. Furthermore, a similarity measure, by integrating the high-level semantic distance and low-level features, is used to retrieve images. The proposed semantic learning scheme provides a way to bridge the gap between the high-level semantic concept and low-level features for content-based image retrieval. Experimental results show that the performance of the proposed method is excellent when compared with that of other methods.

Index Terms—SVM decision boundary; Content-based image retrieval; Principal axis analysis; Semantic learning.

I. INTRODUCTION

As the Internet advances, the demands for storing multimedia information (such as text, image, audio, and video) have increased. With such databases comes the need for a richer set of search facilities that include keywords, sounds, examples, shapes, colors, textures, spatial structures and motion. Traditionally, textual features such as filenames, captions and keywords have been used to annotate and retrieve images. As they are applied to a large database, the use of keywords becomes not only cumbersome but also inadequate to represent image content. Many content-based image retrieval systems have been proposed in the literature [1-2]. Accessing images based on their content is widely used as indexing features for image retrieval. Most of the past research was devoted to the development of an efficient image retrieval system based on the low-level feature. However, the semantic properties of the underlying data are not correctly captured in general. Deriving high-level semantic concepts from low-level visual features is typically based on one of two approaches as: (1) learning from interactive user relevant feedback and (2) learning from templates without run-time user interaction. In the relevance feedback approach [3], one may formulate learning of user perception as an optimization problem to estimate parameters of the distance metric to minimize the sum of distances of relevant examples from the query in the low-level feature space based on user feedback. However, since the set of semantically similar objects might be classified into several clusters in the low-level feature space, querying an object in one cluster would not be able to retrieve semantically similar objects in other clusters by estimating parameters of the distance metric.

Learning from templates, on the other hand, attempt to discover hidden semantics of images based on similarity of low-level visual features [4]. Most image recognition methods focus on certain semantic concepts whose features are highly discriminated for particular user-defined classes. One problem with the learning from templates approach is that the templates were being generated semi-automatically in general. The process to generate meaningful templates for a very large image database is essential but overlooked.

Two important things are needed to address the problem of image interpretation to generate semantic templates which conform to the viewpoint of human vision: (1) some objects in an image are semantically relevant with respect to a query image, the other non-relevant objects can reduce the accuracy of the captured hidden semantic concepts based on the low-level features of a region; (2) besides color, shape and texture, spatial distribution is useful for semantic learning and semantic classification. In this paper, we focus on the problem of semantic space segmentation to improve the performance of content-based image retrieval using support vector machine (SVM) decision boundary and principal axis analysis. In our approach, the query and all the images in a database are segmented into multiple disjoint regions, each of them are represented by three types of low-level features (i.e. color, shape, and texture). First, we proposed a region-based retrieval method to discover hidden semantic regions using the Expectation-Maximization (EM) technique, which are called objects in this paper. The activity scope of hidden semantic objects will be found using the support vector

Manuscript received December 25, 2007.

Tian-Luu Wu is with the Electronic Engineering, National Kinmen Institute of Technology, Taiwan, R.O.C. (corresponding author to provide phone: +886-82-313559; fax: +886-82-313304; e-mail:wtlu@kmit.edu.tw).

Ji-Hwei Horng is with the Electronic Engineering, National Kinmen Institute of Technology, Taiwan, R.O.C. (e-mail:horng@kmit.edu.tw).

machine and the support vector will be decided by the proposed principal axis analysis. Furthermore, a similarity measure, by integrating the high-level semantic distance and low-level features (color, shape, and texture), is used to retrieve images. Experimental results show that the performance of the proposed method is excellent when compared with that of QBIC CBIR system and the Zhang and Zhang's proposed system.

II. The overview of support vector machines

From the viewpoint of image-semantics, each object has its activity scope. Finding the maximum and fair activity scope for every object in an image is the best way to guide hidden semantic concepts. Many existing linear classifiers (hyperplanes) can separate the space of objects. However, only one achieves maximum separation. SVMs were a set of related supervised learning methods for pattern classification and function regression, which was developed by Vapnik [5]. It has also been proved to be very successful in many other applications such as handwritten digit recognition, face detection, and texture classification. A special property of SVMs is known as a maximum margin classifier which can simultaneously minimize the empirical classification error and maximizes the geometric margin.

A. Binary classification

The primary technique of support vector machine (SVM) is to use a high-dimension space to find a hyper-plane to do binary division, where the achieved error rates are minimal. A SVM uses a portion of data to train the system and finds several support vectors that represent training data. These support vectors will be formed into a model and represented in a category using the SVM. According to this model, the SVM will classify a given unknown document by the following classification decision formula:

$$(x_i, y_i),\ldots,(x_n, y_n),\ x \in R^m,\ y \in \{+1,-1\}.$$

where $(x_i,y_i),\ldots,(x_n,y_n)$ are training samples, n is the number of samples, m is the input dimension, and y belongs to category of +1, -1, respectively. The constant y denotes the class to which the point x_i belongs.

If the training data are linearly separable, we can select these hyper-planes so that there are no points between them and then try to maximize their distance. By using geometry, we find the distance between the hyper-planes is $2/|w|$, so we minimize $|w|$. To exclude data points, we need to ensure that i is either:

$$W \cdot X_i - b \geq 1$$

or

$$W \cdot X_i - b \leq -1$$

However, it is not easy to find a hyper-plane to classify the data in many problems. Several kernel functions of SVM can referred to Ref. [8], which applied to solve different problems. Selecting the appropriate kernel function can solve the problem of linear inseparability.

B. Multi-class classifier of SVM

Many real-world classification problems involve more than two classes. Attempts to solve q-class problems with SVM have involved training q SVMs. Each of which separates a single class from all remaining classes, or training $q(q-1)/2$ machines. The first type of classifiers is usually called one-*vs*.-all and classifiers of the second type are called pair-wise classifiers. When the one-vs.-all classifiers are used, a test point is classified into the class whose associated classifier has the highest score among all classifiers. In the case of pair-wise classifiers, a test point is classified in the class which gets the most results from all the possible classifiers [6].

III. Region-Based Image Representation for Hidden Semantic Concept Discovery

The "hidden semantics discovery" concept requires that similarity measures for low-level visual features, such as color, texture, and color-layout, be defined. Ideally, what we try to measure is the semantic similarity which physically is very difficult to define, even to describe. Humans tend to use high-level concepts in everyday life. However, what exiting computer vision techniques can automatically extract from images are mostly low-level features. Object segmentation and recognition is the primary step of computer vision for applying to image retrieval of higher-level image analysis [3]. Automatic segmentation and recognition of objects via object models is a difficult task without a prior knowledge about the shape of objects.

Instead of segmentation and detailed object representation, several image-to-image similarity measurements that combine information from all of regions have been proposed [7]. Although the defined region-to-region similarities attempt to approximate the semantic similarity, the approximation is heuristic and not reliable. Hence, the retrieval accuracy is limited with these systems because not all of regions in an image are semantically relevant with respect to a query image. The irrelevant regions hinder the accuracy of the captured hidden semantic concepts based on the low-level features of regions. To solve the problem, in [8] the authors suggest employing relevance feedback in retrieval to remove the irrelevant regions for improving the retrieval accuracy according to a region-based representation.

Instead of identifying irrelevant regions using relevance feedback in this paper, we propose a region-based retrieval method to discover hidden semantic regions in an image using Expectation-Maximization technique. The images in a database are first segmented into homogenous regions. A kind of color feature, called primitive, will be extracted to discover hidden semantics from each region using the Expectation-Maximization technique. Then a novel method of semantic learning for content-based image retrieval using a decision boundary of support vector machine is proposed, which can decide the activity scope of a region. Last, the multiple low-level features such as color, shape, and texture and semantic-level features are extracted to represent the content of each region for the image retrieval.

A. Image Segmentation and the Primitive Extraction

In this section, we will describe a kind of color feature called primitive which will be used for color image retrieval. The *YIQ* model is used as the color model. The image is first segmented into a disjoint regions, the mean color of region is represented its region. Assume that an image is segmented into n's regions, that are, $R_1, R_2, R_3, \ldots, R_n$, and the region R_i has m's neighbors. Let $f(R_i) = (Y_{R_i}, I_{R_i}, Q_{R_i})$ denote the mean color of region R_i, and the primitive value of color is defined as:

$$P_{R_i} = [P_{R_i}^1, P_{R_i}^2, P_{R_i}^3, P_{R_i}^4, P_{R_i}^5, P_{R_i}^6, \ldots P_{R_i}^{3m+1}, P_{R_i}^{3m+2}, P_{R_i}^{3m+3}] \quad (1)$$

where

$$P_{R_i}^1 = \alpha_1(Y_{Ri} - \bar{Y}), P_{R_i}^2 = \alpha_2(I_{Ri} - \bar{I}), P_{R_i}^3 = \alpha_3(Q_{Ri} - \bar{Q}),$$
$$P_{R_i}^4 = \alpha_1(Y_{R_i} - Y_{R_{j,1}}), P_{R_i}^5 = \alpha_2(I_{R_i} - I_{R_{j,1}}), P_{R_i}^6 = \alpha_3(Q_{R_i} - Q_{R_{j,1}}),$$
$$P_{R_i}^7 = \alpha_1(Y_{R_i} - Y_{R_{j,2}}), P_{R_i}^8 = \alpha_2(I_{R_i} - I_{R_{j,2}}), P_{R_i}^9 = \alpha_3(Q_{R_i} - Q_{R_{j,2}}),$$
$$\vdots$$
$$P_{R_i}^{3m+1} = \alpha_1(Y_{R_i} - Y_{R_{j,m}}), P_{R_i}^{3m+2} = \alpha_2(I_{R_i} - I_{R_{j,m}}), P_{R_i}^{3m+3} = \alpha_3(Q_{R_i} - Q_{R_{j,m}}),$$

and $\alpha_1, \alpha_2,$ and α_3 are the weighting factors of color and set to be 0.4, 0.3, 0.3, respectively in this study. $\bar{Y}, \bar{I}$, and $\bar{Q}$ denote the mean color of whole regions in an image.

The research for the importance of primitives in an image belongs to Gaussian distribution domain [9]. Every region is given a weighting to reveal the dominance of a region using the primitive value of color, which is called weighting of dominant region and defined as follows:

$$RW_i = 1 - e^{-\frac{VP_{R_i}}{S_v}} \quad (2)$$

where s_v denotes the covariance of primitive value of color in an image, and VP_{R_i} is a primitive value of region R_i from (1), which is defined by

$$VP_{R_i} = \sum_{1 \le k \le 3m+3} P_{R_i}^K. \quad (3)$$

The RW_i indicates the weighting of dominant region R_i, which is obtained using the primitive difference from the region R_i and its neighboring regions. An image is segmented into n's non-overlapping regions, that is $R_i, i = 1, \ldots, n$. The arrangement of RWs in ascending order ($RWI_1 \ge RWI_2 \ge \cdots \ge RWI_n$), associates the same ordering to the region R_is

$$R_1 \ge R_2 \ge R_3 \ge \cdots \ge R_n \quad (4)$$

where R_1 is called the most dominant region and R_n is the least dominant region. A ratio is decided to reserve how much of the regions can capture the semantics of an image and the remainder will be ignored. The radio ρ is defined by

$$\rho = \frac{\sum_{1 \le i \le p} RW_i}{\sum_{1 \le j \le n} RW_j} \quad (5)$$

where n denotes the number of regions and p ($p<n$) denotes preceding sequences in eq. (4).

B. Finding Support Vectors using Principal Axis Analysis

SVM performs the semantic space segmentation for two objects by determining the separating hyper-plane with the maximum distance to the closest points training set. These support vectors are decided by the principal axis analysis in this paper, which is used to construct a hyper-plane projection line with minimum inertia. The projection scores, which are obtained by projecting the vectors in two objects on the line, are used as the approximations for filtering. Given a central vector x of an object, we project a set of vectors in the other object and a support vector can be found based on the projection score of x. The algorithm of principal axis analysis has been proved to achieve good results in terms of retrieval accuracy [10].

To find the decision function of SVM, the principal axis must be determined. The principal axis can be conveniently represented in terms of moments. In the case of 2-dimensional feature space S^2, the principal axis L can be represented as

$$\frac{x}{\cos\alpha} = \frac{y}{\cos\beta} \quad (6)$$

Where α and β are the angles between the principal axis L and the x and y, respectively. $Cos\,\alpha$ and $\cos\beta$ are the two direction numbers of L and satisfy the following relationship:

$$\cos^2\alpha + \cos^2\beta = 1 \quad (7)$$

Let $\bar{z} = (\cos\alpha, \cos\beta)$ be a vector aligned with L. The distance from a color vector $\vec{c}$ to L is given by

$$d = \frac{\|\bar{z} \times (\bar{c} - \bar{c})\|}{\|\bar{z}\|} \quad (8)$$

where $\bar{c}$ is the centroid of the feature space and is defined as:

$$(\bar{x}, \bar{y}) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i, \frac{1}{N}\sum_{i=1}^{N} y_i\right) \quad (9)$$

where x_i and y_i are the two-stimulus position values of the ith vector. For processing convenience, we subtract the centroid from all the feature vectors in order to translate the origin of the feature space to the centroid.

To find the direction number of the 2-dimensional principal axis L, use the origin at the centroid. Then the moment of inertia of the point in s^2 about the line L is

$$I(\alpha, \beta) = \sum_{x,y \in s^2} [(x\cos\beta - y\cos\alpha)^2] \quad (10)$$

Differentiate this with respect to $\cos\alpha$ and $\cos\beta$, and equating to zero gives

$$\begin{aligned} m_{0,2}\cos\alpha - m_{1,1}\cos\beta &= 0 \\ -m_{11}\cos\alpha + m_{2,0}\cos\beta &= 0 \end{aligned} \quad (11)$$

Equation (11) is a homogeneous system, which has a trivial solution $\cos\alpha = \cos\beta = 0$. However, this solution does not satisfy the equation (7) and hence the nontrivial solution of system should be found by performing of Guass-Jordan reduction procedure [11] on the augmented matrix [M|0]. The result is

$$\begin{bmatrix} 1 & 0 & k_1 \mid 0 \\ 0 & 1 & k_2 \mid 0 \end{bmatrix}$$

where $k_i, i = 1,2$ are the ratios of $\cos\alpha$ and $\cos\beta$. Based on the reduction results, the values of $\cos\alpha$ and $\cos\beta$ can be computed as

$$(\cos\alpha, \cos\beta) = \left(\frac{k_1}{\sqrt{1+k_1^2+k_2^2}}, \frac{k_2}{\sqrt{1+k_1^2+k_2^2}}\right). \quad (12)$$

Once the principal axis L is obtained, we can project each position vector onto L to compute the projection score of vector by the following equation:

$$P_{\bar{c}} = x\cos\alpha + y\cos\beta \quad (13)$$

Consider two objects in an image which consists of two large numbers of vectors and represented as $A=\{a_1,a_2,\ldots,a_m\}$ and $B=\{b_1,b_2,\ldots,b_n\}$, respectively. The centroid of objects A and B is found as

$$C_{\vec{A}}(\bar{x}_1,\bar{y}_1)=\left(\frac{1}{m}\sum_{i=1}^{m}x_{a_i},\ \sum_{i=1}^{m}y_{a_i}\right)$$
$$C_{\vec{B}}(\bar{x}_2,\bar{y}_2)=\left(\frac{1}{n}\sum_{j=1}^{n}x_{b_j},\ \sum_{j=1}^{m}y_{bj}\right) \tag{14}$$

where $(\bar{x}_1,\bar{y}_1)$ and $(\bar{x}_2,\bar{y}_2)$ denote the vectors of the centroid position of objects A and B, respectively. For processing convenience, each vector in the space is subtracted by the centroid and the results in translating the origin of the feature space of the centroid. The centroid vectors $C_{\vec{A}}(\bar{x}_1,\bar{y}_1)$ and $C_{\vec{B}}(\bar{x}_2,\bar{y}_2)$ were chosen to construct the principal axis L of space. To find the support vectors, the vectors of objects A and B will be projected onto the principal axis. Let the projection scores v_is and u_js associate with the same vectors a_is and b_js, respectively. The projection scores of support vectors will be found as

$$d_{\vec{V}_1}=\min_i(D(C_{\vec{B}},v_i)),\ i=1,\ldots,m,$$
$$d_{\vec{V}_2}=\min_j(D(C_{\vec{A}},u_j)),\ j=1,\ldots,n \tag{15}$$

where $D(\cdot)$ denotes the distance metric using the Euclidean distance. The coordinates of support vector a_i and b_j on the principal axis can be calculated as

$$P(x_{\vec{v}_1},y_{\vec{v}_1})=\left(\frac{d_{\vec{v}_1}\cdot x_{C_{\vec{B}}}+\left(\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}-d_{\vec{v}_1}\right)x_{C_{\vec{A}}}}{\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}},\frac{d_{\vec{v}_1}\cdot y_{C_{\vec{B}}}+\left(\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}-d_{\vec{v}_1}\right)y_{C_{\vec{A}}}}{\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}}\right)$$

$$Q(x_{\vec{v}_2},y_{\vec{v}_2})=\left(\frac{d_{\vec{v}_2}\cdot x_{C_{\vec{A}}}+\left(\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}-d_{\vec{v}_2}\right)x_{C_{\vec{B}}}}{\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}},\frac{d_{\vec{v}_2}\cdot y_{C_{\vec{A}}}+\left(\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}-d_{\vec{v}_2}\right)y_{C_{\vec{B}}}}{\sqrt{(x_{C_{\vec{B}}}-x_{C_{\vec{A}}})^2+(y_{C_{\vec{B}}}-y_{C_{\vec{A}}})^2}}\right) \tag{16}$$

where $d_{\vec{V}_1}$ and $d_{\vec{V}_2}$ denote the projection scores of the support vectors a_i and b_j projected on the principal axis L according to eq. (16). A perpendicular bisector will be found by

$$\left(\cos\alpha-\frac{x_{\vec{v}_1}+x_{\vec{v}_2}}{2},\cos\beta-\frac{y_{\vec{v}_1}+y_{\vec{v}_2}}{2}\right)\cdot(x_{\vec{v}_1}-x_{\vec{v}_2},y_{\vec{v}_1}-y_{\vec{v}_2})=0. \tag{17}$$

The decision boundary of SVM is simple to obtain by

$$d(\cos\alpha,\cos\beta)=(x_{\vec{v}_1}-x_{\vec{v}_2})\cos\alpha+(x_{\vec{v}_1}-x_{\vec{v}_2})\cos\beta-\frac{(x_{\vec{v}_1}^2-x_{\vec{v}_2}^2)+(y_{\vec{v}_1}^2-y_{\vec{v}_2}^2)}{2}. \tag{19}$$

The support vector machine will be constructed using two classes of objects shown as Fig. 1.

C. Multi-class classifier of SVM

An image usually uses more than two objects to represent the semantic of the image. The proposed scheme based on pair-wise classifiers is used to solve the many classifier problems. To improve the effectiveness of the proposed method, the intermediate classification strategies in the style of ECC [12] is applied in the proposed method. The completing of proposed classification algorithm based on the multi-class classifier of SVM is given as follows.

Algorithm 1: Proposed classification algorithm based on the multi-class classifier of SVM.
Input: A set of objects which represents the semantic of an image.
Output: A set of decision boundary function F_i which separate the semantic space of an image.
Method:

1. Let R denote the collection of object sequences ($r_1,r_2,\ldots,r_n$) of an image, where each of them is constructed from the scanning of an image from left-to-right and top-to-bottom.
2. Let D and T denote a collection consisting of the object selected and empty initially.
3. Select an object r_i from R.
4. While ($R\neq$ NULL)
   4.1. $C_R=r_i$.
   4.2. Let R_n denote the neighbor objects of C_R in R.
   4.3. Find the centroid of object r_i using eq. (14).
      4.3.1. While ($R_N \neq NULL$)
      4.3.2. Select the object r_j from R, which is the neighbor object of the objects r_i in R.
      4.3.3. Find the centroid of object r_j using eq. (14).
      4.3.4. Construct the principal axis L using the centroid of objects r_i and r_j.
      4.3.5. Project the vectors of each object onto the principal axis using eq. (13).
      4.3.6. Find the projection position of support vector S_A and S_B using the eq. (15).
      4.3.7. Find the decision boundary function F_i of SVM using the eq. (17).
      4.3.8. Remove the selected object r_j from R, and add it to T.
   4.4. Remove the selected object r_i from R, and add it to D.
   4.5. Remove the object in T, and add it to R.
   4.6. Select an object r_i from R.
5. Return the set of decision boundary function F_i.

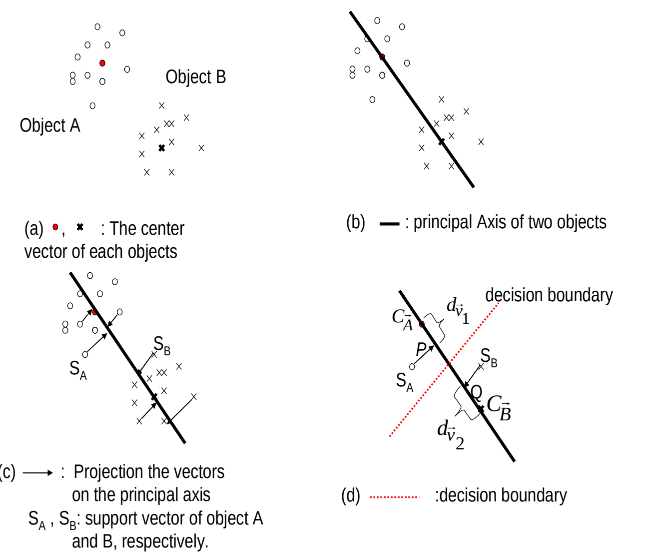

Fig.1. An example illustrates that the decision boundary will be decided using the support vector: (a) two objects are reserved to present the semantic of an image; (b) the principal axis is found using the centroid vectors connection;

(c) projection of the vectors of each objects on the principal axis; (d) the decision boundary is decided by using eq.(18).

IV. IMAGE RETRIEVAL STRATEGIES

Three types of low-level features, i.e., color, shape, and texture are extracted from the objects of each database image. According to the QBIC method [13], the three low-level features are the color histogram, digital central moments and triple (contrast, coarseness, directionally) for color features, shape feature, and texture feature, respectively. In the QBIC system, three features are used independently. Users are required to set the weighting of each features type in order to retrieve semantically relevant images. Unfortunately, the low-level features and the high-level semantic concepts do not have an intuitive relationship. This limits the retrieval accuracy of QBIC-like systems. In this paper, we proposed a method to integrate the low-level features and semantic-level features. Given a query image with the object vector $I_q = (O_1^{(q)}, O_2^{(q)}, ..., O_n^{(q)})$ and an object vector $I_k = (O_1^{(k)}, O_2^{(k)}, ..., O_m^{(k)})$ belonging to database image, two kinds of semantic-level features are used as high-level semantic features. They are the dominant values *RW* of object defined as eq. (2) and its activity scope *S* of the dominant object belonging to its image. Without loss of generality, we assume the number of objects of a database image is larger than that of a query image, i.e. n<m. In order to calculate the value of similarity between two images I_q and I_k, two distances of matching $d_m^{(S)}$ and un-matching $d_u^{(S)}$ on the basis of semantic-level are computed by

$$d_m^{(S)} = \sum_{i=1}^{n}(w_{ij} \| RW_i^{(q)} - RW_j^{(k)} \| + u_{ij} \| S_i^{(q)} - S_j^{(k)} \|) \quad (19)$$

$$d_u^{(S)} = \frac{1}{m-n} \sum_{j=n-m}^{m} \Pi_j \quad (20)$$

where w_{ij} and u_{ij} denote weighting factors of the dominant value of an object and the activity scope of an object and set to be 0.6 and 0.4 respectively. The activity scope S_i of an object and the penalty Π of un-match object is given as

$$S_i = 1 - \frac{Activity\ Scope\ of\ Object_i}{Size\ of\ an\ Image} \quad (21)$$

$$\Pi = \| Object_j\ \ of\ \ Image_{I_k} - Background\ \ of\ \ Image_{I_k} \| \quad (22)$$

where the $background\ \ of\ \ Image_{I_K}$ is the neighbor and omitted with the $object_j\ \ of\ \ Image_{I_K}$. Note that the values of $d_m^{(S)}$ and $d_u^{(S)}$ should be normalized by maximum values of $d_m^{(S)}$ and $d_u^{(S)}$, respectively, before advancing to define similarity measure between the two images I_q and I_k. This normalization process will result in the normalized values of $d_m^{(S)}$ and $d_u^{(S)}$ are confined within the (0, 1) interval. Let $\tilde{d}_m^{(S)}$ and $\tilde{d}_u^{(S)}$ denote the normalized $d_m^{(S)}$ and $d_u^{(S)}$, respectively. The measurement of similarity $d^{(S)}$ between two images I_q and I_k on the basis of high-level of semantic distance is defined as

$$d^{(S)} = 1 - (w_1 \tilde{d}_m^{(S)} + w_2 \tilde{d}_u^{(S)}) \quad (23)$$

where w_1 and w_2 represent the weighting of $\tilde{d}_m^{(S)}$ and $\tilde{d}_u^{(S)}$ respectively, and $w_1+w_2=1$.

The low-level feature can be calculate using the weighted Euclidean distance as

$$d^{\Gamma} = \lambda_1 \| I_q^C - I_k^C \| + \lambda_2 \| I_q^S - \vec{I}_k^S \| + \lambda_3 \| I_q^T - I_k^T \| \quad (24)$$

where d^{Γ} denote the low-level features distance and λ_1, λ_2, and λ_3 are the weighting factors of the low-level features color, shape, and texture, and set 0.4, 0.3, and 0.3, respectively. The value of d^{Γ} is also normalized by the maximum value of d^{Γ}, and defines as

$$d^{(L)} = 1 - \frac{d^{\Gamma}}{d_{\max}^{\Gamma}} \quad (25)$$

The overall distance between the two models is a weighting sum of the semantic distance and the low-level feature distance

$$d^{overall} = \rho^{(S)} d^{(S)} + \rho^{(L)} d^{(L)} \quad (26)$$

where $\rho^{(S)}$ and $\rho^{(L)}$ are the semantic weighting and the low-level feature weighting, respectively, and $\rho^{(S)} + \rho^{(L)} = 1$.

V. EXPERIMENT RESULTS

In order to evaluate the performance of the proposed approach, a series of experiments were conducted on an Intel PENTIUM-IV 2.5GHz PC. The Zhang and Chen's method [14], R. Zhang and Z. Zhang [8], and QBIC method [13] are also simulated by computer software for the purpose of performance comparison. A real data set consisting of 16,887 natural images were used for the test database. The image database has images of 8,956 scenes, 1,189 persons, 1,303 animals, 1,584 plants, 1,856 traffic tools, 389 buildings, 270 sports, 565 pictures, 476 arts, and 299 others. Each image in the real database is first tailored to the size of 256×256 for testing the retrieval approach. Ten percent of the images from the different category of image database are randomly chosen to decide the ratio of dominant region in eq. (5). Table 1 shows the ratio of capturing the semantic of an image with different category of images using the proposed method. According to simulation results, the proposed method provides a dramatic result when the ratio of reserving regions was assigned to be larger than 0.85.

It is difficult to derive a formal method in evaluating the retrieval accuracy of the tested database system. Traditional metrics for evaluating performance are recall and precision. They are a function of both correct matches and relevance of a database image to a query. The retrieval accuracy measured by precision and recall is computed as follows. Recall measures the ability of the system to retrieve all images that are relevant and defined as:

$$\text{Recall} = \frac{relevances\ correctly\ retrieved}{all\ \ relevances}.$$

Precision measures the ability of the system to retrieve only images that are relevant and can be computed by:

$$\text{Precision} = \frac{relevances\ correctly\ retrieved}{all\ \ retrieved}.$$

Recall and precision required a ground truth to assess the relevance of images for a set of significant queries.

ISBN: 978-988-98671-8-8

The average precision and recall curves are plotted in Figs. 2 and 3, respectively. It can be seen that the proposed method achieves good results in terms of retrieval accuracy compared with Zhang and Chen's method [14], R. Zhang and Z. Zhang's method [8], and QBIC method [13]. The performance of the proposed method provides a good way to evaluate the fitness of a semantic description used to retrieve the semantic content of the image.

Table 1. The results of the average number for the ratio of capturing the semantic of an image using the different category of database images.

The category of image	The ratio of capturing the semantic of an image (%)			
	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.85$	$\rho = 0.9$
scenes	46	80	85	86
persons	60	84	90	87
animals	42	84	90	88
plants	45	80	91	92
traffics	44	79	90	88
buildings	45	75	90	90
sports	55	78	91	92
pictures	45	80	93	89
arts	45	82	92	90
others	48	84	90	90

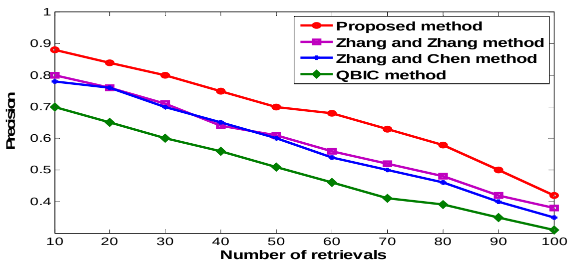

Fig.2. Average precision versus number of retrievals.

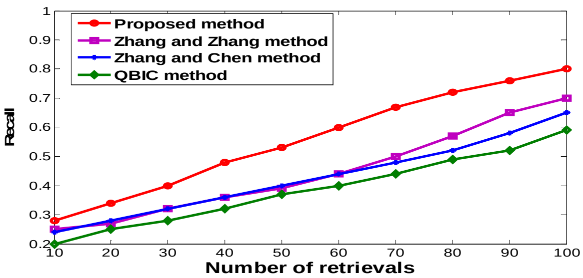

Fig.3. Average recall versus number of retrievals.

VI. Conclusions

In this paper, we have presented a new method of semantic space segmentation for content-based image retrieval using SVM decision boundary. Moreover, special attention is paid to solve the problem of SVM decision boundary by principal axis analysis. The similarity measurement based on the proposed semantic distance metric has been defined to capture the perceptual and semantic meanings. Compared with Zhang and Zhang's method, Zhang and Chen's method, and QBIC method, the experimental results demonstrate that the proposed method outperforms other methods in terms of retrieval accuracy in the image database, which shortens the gap between the low-level feature and high-level semantic access methods. Some possible improvements and future research topics are as follows.

Firstly, the proposed method can be extended to describe the semantic of images when there is a better way is provided to determine the threshold. Furthermore, the proposed method can also be useful for applications of semantic classification and image annotation.

References

1. Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, "A survey of Content-based image retrieval with high-level semantics," Pattern Recognition, vol.40, pp262-282, 2007.
2. A. Mojsilovic, B. Rogowitz, "ISee: perceptual features for image library navigation," Proc. Of the SPIE, Human Vision and Electronic Image, Vol. 4462, pp266-277, 2002.
3. Y. Rui and T. S. Huang, "Optimizing Learning in image retrieval," in Proc. Of IEEE Int. CVPR, Jun. 2000.(3)
4. Z. Wang, Z. Chi, D. Feng, and A. C. Tsoi," Content-based image retrieval with relevance feedback using adaptive processing of tree-structure image representation, Int. J. Image Graphics, 3(1), pp. 119-144, 2003.
5. V. Vapnik," The nature of statistical learning theory," New York: Springer-verlag; 1995.
6. H.-C. Kim, S. Pang, H.-M. Je, D. Kim, S. Yang, and Sung, "Constructing support vector machine ensemble," Pattern Recognition, 36(12), pp. 2757-2767, Dec. 2003.
7. A. W. M. Smeulders, M. Worring, A. Gupta, and R. Jain, "Content-based image retrieval at the end of early years," IEEE Trans. Pattern. Anal. Machine Intell., 22(12), pp 1349-1380, 2000.
8. R. Zhang and Z. Zhang, Hidden semantic concept discovery in region based image retrieval," in Proc. Of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'04), 2004.
9. M. Karger and S. Sandomirsky, "Multidimensional statistical technique for detection of low contrast geochemical anomalies," J. of geochemical Exploration, 72, pp.47-58, 2001.
10. S. C. Cheng and T. L. Wu, "Subpixel edge detection of color images by principal axis analysis," Pattern Recognition, 38, pp. 527-537, 2005.
11. B. Kolman, Introductory linear algebra with applications, Prentice-Hall, Inc. 1997.
12. Y. Yao, G. Y. Marcialis, M. Pontil, P. Frasconi, and F. Roli, " Combining flat and structured representations for fingerprint classification with recursive neural networks and support vector machines," Pattern Recognition, 36(2), pp. 397-406, 2006.
13. M. flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker," Query by image and video content: the QBIC system, IEEE Computer, 28(9), pp23-32, 1995.
14. C. Zhang, and T. Cheng,"An Active Learning Framework for Content-based Information retrieval," IEEE trans. On Multumedia, 4(2), pp.260-268, June 2002.